Occupancy Modeling Case Study #2: Factors Influencing Large Scale Bobcat Distribution in the Eastern United States
SMSC Terrestrial Mammals Course 2016

Project Description and Context:

In this module, we will fit models that involve covariates for both occupancy and detection probabilities to data that have been collected with camera traps by citizen volunteers across 7 eastern US states including Virginia and Washington DC. At each study location, 27-80 camera sites were established either on or off hiking/wildlife trails in 32 different state and regional parks, national forests and small reserves. The goal of this project case study is to describe the regional distribution of the bobcat (*Lynx rufus*), investigating which large scale factors are most important. We will use occupancy modeling to investigate the influence of 4 different site covariates on probability of occupancy: 1) level of human development, 2) amount of human use of the site, 3) regional location (i.e. latitude and longitude), and 4) vegetative productivity. Surrounding human development was recorded as the housing density within 5km of each camera site. Human site use was recorded as the # of human photos recorded at the camera station. The site's position in the regional landscape was represented by its latitude and longitude, and vegetation productivity was based on remotely sensed data and represented by the Normalized Difference Vegetation Index (NDVI). An additional row of information includes an ID code for each park. In addition, we have two additional covariates that could potentially influence the probability of detection at each site. The maximum detection distance of each camera was recorded as well as whether it was located on a trail.

The data are included in the sample data folder on course's Dropbox Site in the Kolowski>Data folder. There are two files, one representing the detection history at each site (*Bobcat-3day.csv*) and one including the site covariates (*site_covar_3Day_class.csv*). The detection history was created by pooling every 3 days of camera data into 1 occasion, resulting in 7 survey occasions from a 21-day sampling period at each location. There is a total of 1839 camera sites.

Thanks to Tavis Forrester for sharing these data, and see Rota et al. 2016 (Ecology) for more details on one application of these data.

Exercise Objectives

- Practice how to create and run occupancy models where occupancy and/or detection is a function of site specific covariates
- Learn to import data directly as a csv file
- Continue to increase comfort level and familiarity with all aspects of analysis in PRESENCE from data exploration to model selection, and interpretation of results

Data Files: *Bobcat-3day.csv*
site_covar_3Day_class.csv

INSTRUCTIONS

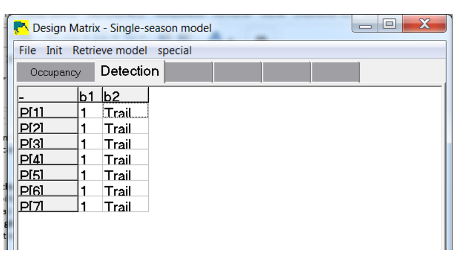
Step 1 – Data Import: Begin PRESENCE, start a new project and name it something like "Bobcat". Open the Input Data Form by clicking the button. Instead of copying and pasting the detection data into Presence (as in the Weta example), we will try a different way (because there are too many rows to copy to the clipboard). First click in the first data cell (Site 1, occasion 1-1). Then Click **Edit>Insert from CSV file** and select your detection data file (it may take a minute or two to load). Double check that you have the right number of rows. The # of surveys/season at the top should be updated to read 7.

Next we have to import our 9 site-specific covariates. Note that our categorical variable representing whether a camera is on or off a trail is represented by two columns (dummy variables) so our total count is 9. Therefore, change the number of site-specific covariates at the top (#site covs) to 9. To input the site-specific covariates, make sure the covariates sheet is highlighted in Presence, and that you are clicked in the first data cell (cell should be a dotted line, not a flashing cursor). Click **Edit>Insert from CSV file** and selected the csv file with your covariates.

Now select **File>Save As** and name your data file, and choose the location to store it (I suggest your folder on the computer desktop). Then close the data input form which will return you to the Initial setup screen. Make sure the data file you just created in listed in the **Enter data filename** window (should happen automatically) and click **OK.** You'll be asked to agree to two windows, telling you where your project will be stored, and what associated files are being created. After a couple of seconds a blank results browser should appear. If you do not see the results browser, you have not successfully set up your project file.

Step 2 – Running a simple model: First, let's fit a simple model where the probability of occupancy is the same for all sites and the probability of detection is the same in all surveys, and at all sites. In the general notation we could call this model **psi(.),p(.)**. Click **Run>Analysis: Single Season>Simple Single Season** to set up the model. The design matrix for occupancy will be just a '1' in the single grid cell, and for detection probability we have seven rows (for the 7 possible survey occasions) and require a single column of '1's. This is actually the default model that appears when you select to run a custom, single-season model, so after the design matrix appears you just need to hit **'Ok to Run'** then confirm the results to add them to the results browser. You can then view the summary results in the results browser and you can hit the button to view the full results of the model in Notepad.

Step 3 – Analyze the influence of covariates on occupancy and detection: Next we shall fit a model where the probability of detection will vary among plots depending on whether the camera was setup on a trail or off-trail. We could call this model **psi(.),p(trail)**. Start the analysis by selecting **Run>Analysis: Single Season>Simple Single Season** This will bring up the design matrices window. Set the design matrix for detection to look like the following.



This will require you to add a column to the design matrix, then enter a covariate name. To add a column, you simply right-click anywhere on the design matrix to open a pop-up menu; select **Add Col** (the 5th item in the menu) which will create an empty column on the right-hand side of the design matrix. You can then type the name of the covariate into the grid cell, but if you do so you must get the name 100% correct (and note that PRESENCE is case sensitive). Alternatively, if you open the **Init** menu you will notice that all of the available covariates are listed there, prefixed with a '*'. After activating column 2 (so the grid cell in the second column has the dotted outline), select **Init>*Trail** and the covariate name **Trail** will be inserted into the correct column.

You can leave the occupancy matrix as it is. Once the design matrices are set up, return to the SNER window (without closing the design matrix window), <u>rename the model</u> then hit 'OK to Run' and confirm the results when prompted. What this model allows is for detection probability to have a different value for sites on trails, compared to sites off-trails. The first column in the design matrix represents a baseline detection value, and the second column represents the change in that detection when sites are on trails. You'll see this in the detailed results output, where now there are two values listed for *p*.

Note: You can retrieve previously created models and their design matrices so you don't have to start over with every new model. From the design matrix screen click Retrieve Model and select the one you want to retrieve.

EXERCISE:

Working in small groups, continue to investigate your other covariates until you identify the best fitting model that represents these data. First, establish what the best model is that represents detection probability by running and comparing the following models:

- psi(.) p(.)
- psi(.) p(trail)
- psi(.) p(trail + Det_dist)

After deciding on the best detection model, use that to test these remaining models which contain different combinations of covariates affecting occupancy.

- psi(NDVI) p(best model)
- psi(HDens) p(best model)
- psi(People) p(best model)
- psi(NDVI + HDens) p(best model)
- psi(NDVI + People) p(best model)
- psi(HDens + People) p(best model)
- psi(NDVI + People+ HDens) p(best model)

Once you've used AIC values and model weights to determine the most appropriate model, answer the following questions:

1. What is the naïve occupancy estimate for bobcats across the study area?

2. Using the most basic model [psi(.)p(.)], what is the detection probability and estimated probability of bobcat occupancy across all sites?

3. Is there evidence to indicate that detection probability is influenced by whether the cameras were placed on trails? If so, was detection probability higher or lower on trails? Was there an influence of camera detection distance on probability of detection as well? If so, what was it?

4. To review the relative importance of the 4 tested covariates on occupancy, list the summed AIC weights of all 4 variables.

5. What model best represents bobcat occupancy on the landscape? List the important variables, the direction of influence and explain, in an ecological sense, why they might be important in determining large scale bobcat presence on the landscape.

6. How many sample covariates are available with this dataset?

EXTRA WORK:

7. Using all 7 of your models, calculate a model averaged estimate for probability of occupancy of Bobcats across the landscape. (You'll need to paste the results into an Excel file for this).

8. Try to investigate the influence of latitude and/or longitude on bobcat occupancy probability. Is either variable important? If so, why could this be?

9. Investigate the potential impact of human presence at camera stations on our ability to DETECT bobcats in the first place and discuss why someone might use this variable as a covariate for detection.